

COURSERA CAPSTONE PROJECT

Opening a restaurant in Bangalore, India

By

Anand Kumar Singh

Introduction

Bangalore is the capital of the Indian state of Karnataka. It has a population of about 10 million and a metropolitan population of about 8.52 million, making it the third most populous city and fifth most populous urban agglomeration in India. Located in southern India on the Deccan Plateau, at a height of over 900 m (3,000 ft) above sea level, Bangalore is known for its pleasant climate throughout the year. Its elevation is the highest among the major cities of India.

There are many Restaurants in Bangalore, Because of the IT hub of India this city have best in class restaurant across the city, This project is aim to open a new Indian restaurant in Bangalore which will have home like food. Now This city have many good Indian restaurants so finding a neighborhood which have less competition and more profit is the aim of this project.

Business Problem

The objective of this capstone project is to analyse and select the best location in Bangalore city to open a new Indian restaurant. Using Data science methodology and machine learning techniques like clustering, This project provide the solution to the business question: If anyone want to open a good Indian restaurant in Bangalore where they should open it ?

Target Audience

This project is for investors and business persons who wants to open a new Indian restaurant in Bangalore city

Data

To solve this problem

Data required are

- List of neighbourhoods in Bangalore.
- Location of these neighbourhoods in terms of Latitude and Longitude
- Venue data of each neighbourhoods, which can be found from foursquare Api

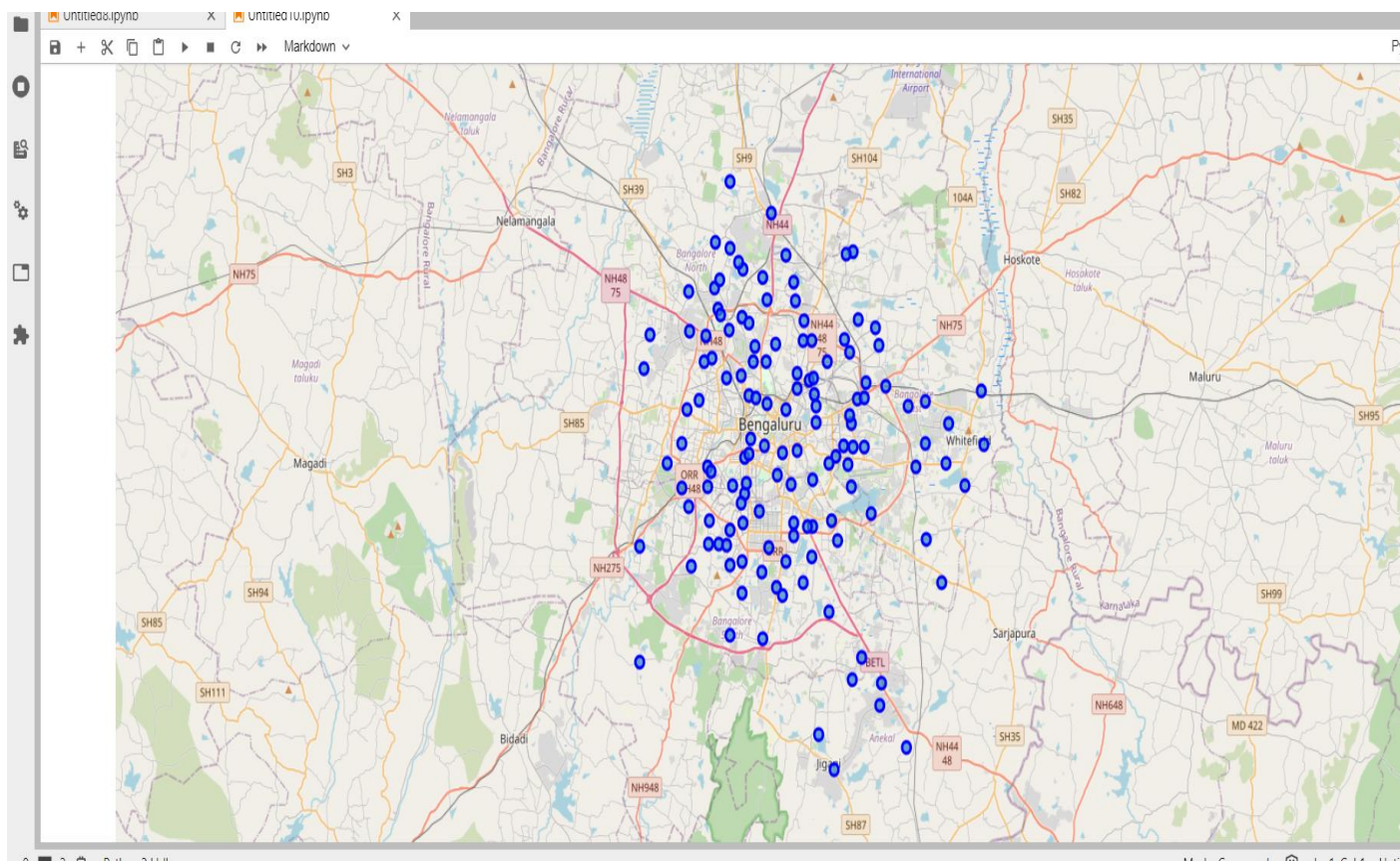
The neighbourhood data is scrapped from Wikipedia page

Link: https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore

The latitude and longitude data is find through geopy library

Neighbourhood of Bangalore

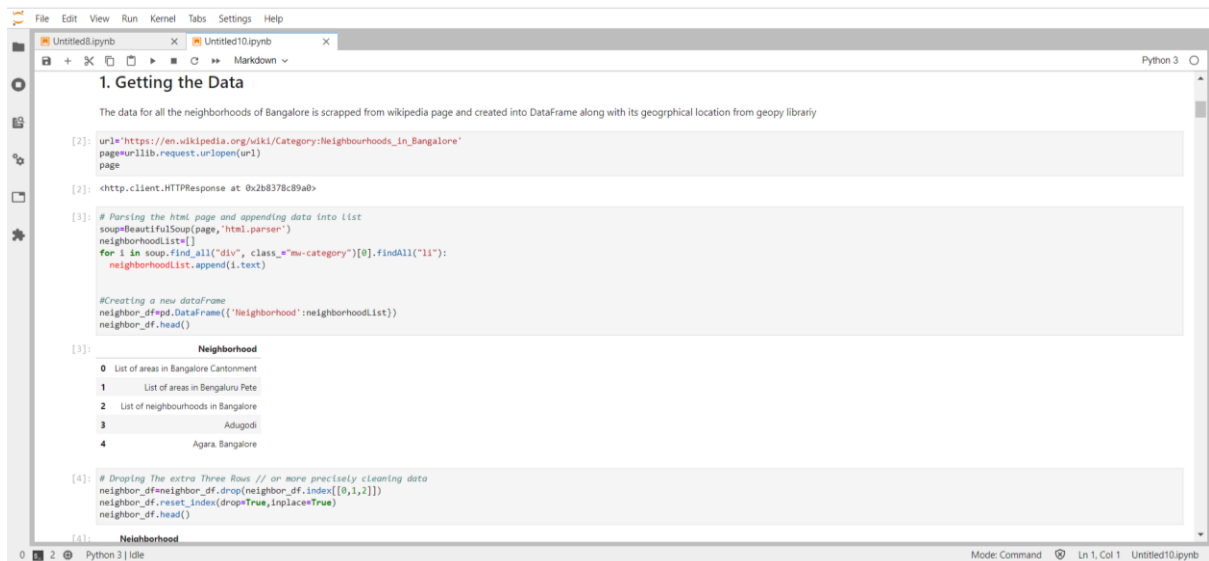
There are 136 neighbourhood in Bangalore city, all the neighbourhood are shown using Folium map , which clearly visualize that where all the major neighbourhoods lies. Creating folium map gives a brief view about all the data points. Now these neighbourhoods will be clustered according to the business problem with the help of KMeans clustering.



Methodology

First we need to collect all the neighbourhoods in Bangalore city, these data is available on Wikipedia page

We do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data, and then create the data frame of it.



```
1. Getting the Data

The data for all the neighborhoods of Bangalore is scrapped from wikipedia page and created into DataFrame along with its geographical location from geopy library

[2]: url='https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore'
page=requests.get(url)

[2]: <http.client.HTTPResponse at 0x2b8378c89a0>

[3]: # Parsing the html page and appending data into list
soup=BeautifulSoup(page,'html.parser')
neighborhoodList=[]
for i in soup.find_all('div', class_='mw-category')[0].findAll('li'):
    neighborhoodList.append(i.text)

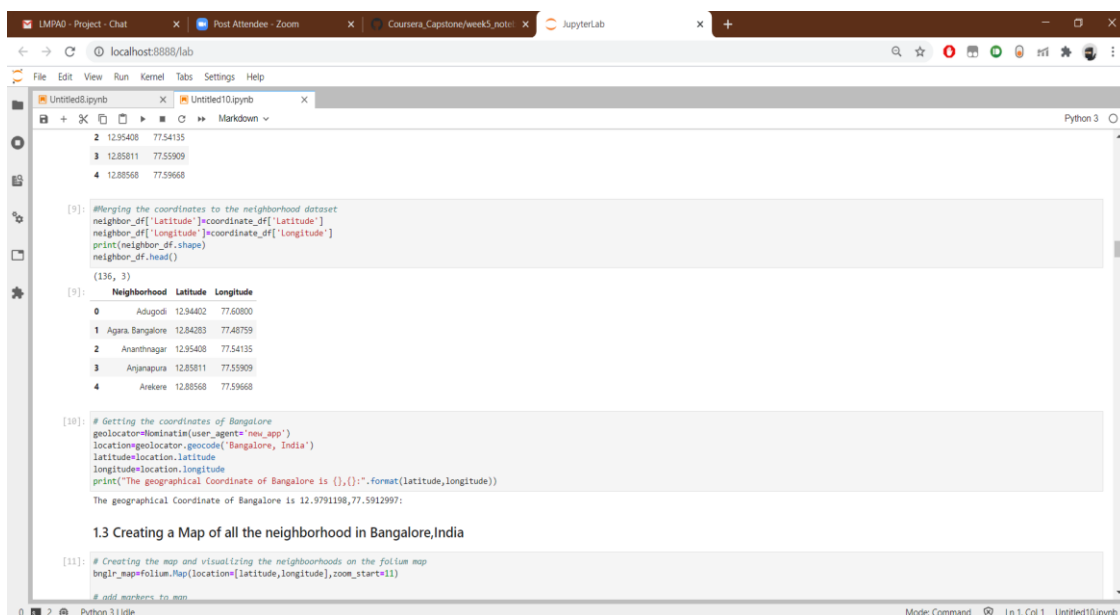
#Creating a new dataframe
neighbor_df=pd.DataFrame({'Neighborhood':neighborhoodList})
neighbor_df.head()

[3]:
Neighborhood
0 List of areas in Bangalore Cantonment
1 List of areas in Bengaluru Pete
2 List of neighborhoods in Bangalore
3 Adugodi
4 Agara, Bangalore

[4]: # Dropping The extra Three Rows // or more precisely cleaning data
neighbor_df=neighbor_df.drop(neighbor_df.index[[0,1,2]])
neighbor_df.reset_index(drop=True,inplace=True)
neighbor_df.head()

[4]:
Neighborhood
0 Adugodi
1 Agara, Bangalore
2 Ananthnagar
3 Anjanapura
4 Ankere
```

Now we have done creating the data frame we need to have all the location of the neighbourhood which is done by geocoder package which allow us to convert the address into geographical coordinates in the form of latitude and longitude in able to use the foursquare api.



```
[9]: #Merging the coordinates to the neighborhood dataset
neighbor_df['Latitude']=coordinate_df['Latitude']
neighbor_df['Longitude']=coordinate_df['Longitude']
print(neighbor_df.shape)
neighbor_df.head()

(136, 3)

[9]:
Neighborhood Latitude Longitude
0 Adugodi 12.84402 77.60800
1 Agara, Bangalore 12.84283 77.48759
2 Ananthnagar 12.95408 77.54135
3 Anjanapura 12.85811 77.55909
4 Ankere 12.85568 77.59668

[10]: # Getting the coordinates of Bangalore
geolocator=Nominatein(user_agent='new_app')
location=geolocator.geocode('Bangalore, India')
latitude=location.latitude
longitude=location.longitude
print('The geographical Coordinate of Bangalore is {},{}'.format(latitude,longitude))

The geographical Coordinate of Bangalore is 12.9791198,77.5912997:

1.3 Creating a Map of all the neighborhood in Bangalore,India

[11]: # Creating the map and visualizing the neighborhoods on the folium map
bnglr_map=folium.Map(location=[latitude,longitude],zoom_start=11)

# add markers to map
```

Next we will use foursquare api to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer account in order to obtain the foursquare id and foursquare secret key. We then make api calls to foursquare passing geographical coordinates of the neighborhoods in python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from the returned data.

```

Untitled8.ipynb x Untitled10.ipynb
Python 3

Yeswanthpur 13 13 13 13 13 13

135 rows x 6 columns

2.2 checking the total number of unique categories

[16]: print('There are {} unique categories.'.format(len(venues_df['VenueCategory'].unique())))
      There are 234 unique categories.

[17]: venues_df['VenueCategory'].unique()[:50]

[17]: array(['Movie Theater', 'Clothing Store', 'Creperie', 'Lounge',
       'Financial or Legal Service', 'Burger Joint', 'Indian Restaurant',
       'Brewery', 'Multiplex', 'Breakfast Spot', 'Café', 'Tea Room',
       'Bar', 'Mobile Phone Shop', 'Dessert Shop', 'Donut Shop', 'Gym',
       'Shopping Mall', 'Coffee Shop', 'Juice Bar', 'Ice Cream Shop',
       'Chinese Restaurant', 'Bakery', 'Gaming Cafe', 'Yoga Studio',
       'Smoke Shop', 'Men's Store', 'Italian Restaurant', 'Arcade',
       'Indian Sweet Shop', 'Bookstore', 'Kerala Restaurant',
       'Snack Place', 'Andhra Restaurant', 'Pizza Place',
       'Mexican Restaurant', 'Fast Food Restaurant', 'Cosmetics Shop',
       'Punjabi Restaurant', 'Boozing Alley', 'Comfort Food Restaurant',
       'Persian Restaurant', 'Eastern European Restaurant',
       'Tibetan Restaurant', 'Thai Restaurant',
       'Paper / Office Supplies Store', 'Seafood Restaurant',
       'Restaurant', 'Sandwich Place', 'Liquor Store'], dtype=object)

3. Analyzing each neighborhoods

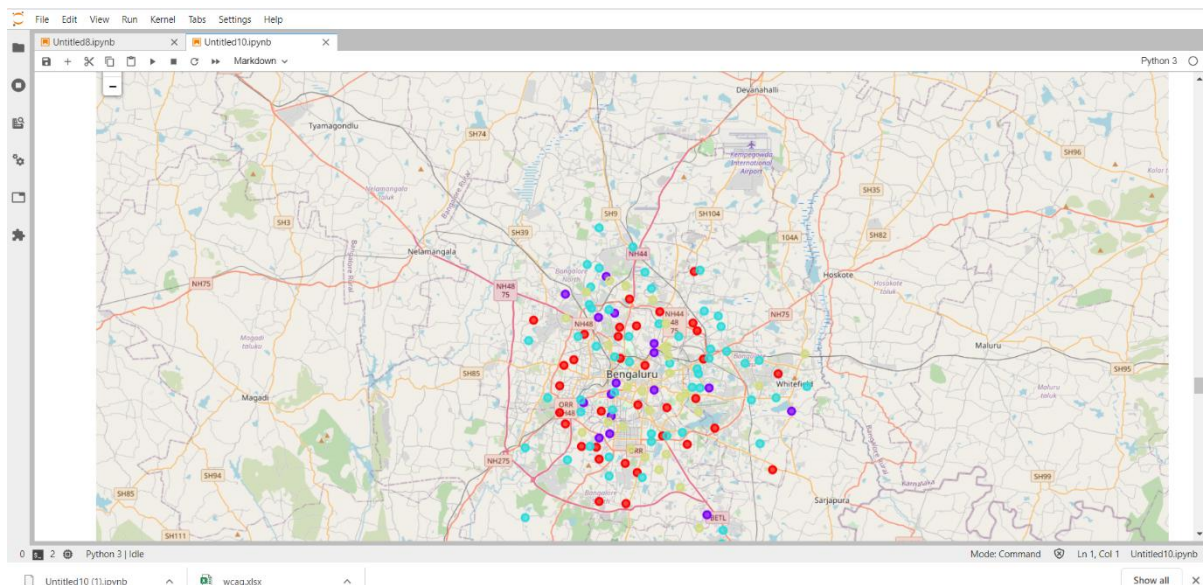
[18]: # one hot encoding
      bangalore_onehot = pd.get_dummies(venues_df[['VenueCategory']], prefix="", prefix_sep="")

      # add neighborhood column back to dataframe
      bangalore_onehot['neighborhoods'] = venues_df['neighborhood']

```

Then we will analyse each neighborhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing data for clustering

And then we will cluster the data and show the resulting cluster on the folium map, this clusters will divide our “Indian Restaurant” to clusters so we can deduce our result.



Here above we can see we have clustered these data into 4 clusters so that it will be easy to determine the result.

Result

The results from the KMeans clustering shows that we can categorize the neighbourhood into 4 clusters based on frequency of occurrence of Indian restaurant.

Cluster 0: Neighborhood with high number of Indian restaurant 34

Cluster 1:Neighbourhood with very less number of Indian restaurant 17

Cluster 2: Neighbourhood with highest number of Indian restaurant 56

Cluster 3: Neighbourhood with moderate number of Indian restaurant 28

Discussion

An observations of from the map and clusters shows that the most of the Indian restaurants are situated in Central Bangalore area. By analyzing each cluster it gives the output that cluster 0 have 34 Indian restaurant across the city, and cluster2 being the most in number i.e 54 Indian restaurants, So This projects states that to get out the maximum profit and getting the best result to open a new Indian restaurant someone should consider opening it in cluster1 area because it has the lowest numbers of restaurant in that area.